U. S. NAVAL SUBMARINE MEDICAL CENTER

Submarine Base, Groton, Conn.

REPORT NUMBER 672

STUDIES IN NAVY COMMUNICATION:

THE EFFECT OF WORD PREDICTABILITY ON SENTENCE INTELLIGIBILITY

by

Joseph R. Duffy
and
Thomas G. Giolas
with
Russell L. Sergeant

Bureau of Medicine and Surgery, Navy Department
Research Work Unit M4306.03-2020DAC5.06

Released by:

J. E. Stark, CAPT MC USN
COMMANDING OFFICER
Naval Submarine Medical Center

8 July 1971

14

Approved for public release; distribution unlimited.

UNCLASSIFIED

Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author)	2a. REPORT SECURITY CLASSIFICATION
NAVAL SUBMARINE MEDICAL RESEARCH LABORATORY, NAVAL SUBMARINE MEDICAL CENTER	UNCLASSIFIED
	2b. GROUP

3. REPORT TITLE

Studies in Navy Communication: The Effect of Word Predictability on Sentence Intelligibility

4. DESCRIPTIVE NOTES *(Type of report and inclusive dates)*

Interim Report

5. AUTHOR(S) *(First name, middle initial, last name)*

Joseph R. Duffy
Thomas G. Giolas
Russell L. Sergeant

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
8 July 1971	8	12

8a. CONTRACT OR GRANT NO	9a. ORIGINATOR'S REPORT NUMBER(S)
N00014-68-A-0197-0001	NSMRL Report No. 672
b. PROJECT NO. M4306.03-2020D.06	
c.	9b. OTHER REPORT NO(S) *(Any other numbers that may be assigned this report)* UCONN IUVC/ONR Report #6
d.	

10. DISTRIBUTION STATEMENT

Approved for public release; distribution unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY	
	NAVSUBMEDRSCHLAB NAVSUBMEDCEN, SUBASE NLON Groton, Conn. 06340	ONR Washington, D.C.

13. ABSTRACT

The primary purpose of this study was to investigate the relationship between word predictability and sentence intelligibility. This relationship was examined by comparing the accuracy of responses by listeners to several lists of sentences. Three methods of scoring used different groups of key words which had previously been judged to represent different degrees of predictability. It was hypothesized that the scores obtained would be a function of the predictability status of the key words used in scoring. Results indicated significant differences between the three scoring procedures for each sentence list over both filtering conditions, and these differences were in the hypothesized direction. The results suggest that the use of easy-to-predict words for scoring purposes will increase sentence intelligibility while the use of difficult-to-predict words will tend to depress the intelligibility of sentences. It was concluded that word predictability is a factor influencing sentence intelligibility and that careful selection of key words for scoring purposes, determined on the basis of their predictability status, may be a possible way of controlling the intelligibility of sentences.

DD FORM 1 NOV 65 1473 (PAGE 1)

S/N 0102-014-6600

UNCLASSIFIED

Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Speech Discrimination Testing						
Speech Intelligibility						
Word Predictability						

STUDIES IN NAVY COMMUNICATIONS:

THE EFFECT OF WORD PREDICTABILITY ON SENTENCE INTELLIGIBILITY

by
Joseph R. Duffy, M.A.
and
Thomas G. Giolas, Ph.D.
University of Connecticut

with
Russell L. Sergeant, Ph.D.
Naval Submarine Medical Research Laboratory

SUBMARINE MEDICAL RESEARCH LABORATORY
NAVAL SUBMARINE MEDICAL CENTER REPORT NO. 672

Bureau of Medicine and Surgery, Navy Development
Research Work Unit M4306.03-2020DAC5.06

Transmitted By:

J. Donald Harris

J. Donald Harris, Ph.D.
Head, Auditory Research Branch

Reviewed and Approved by:

Charles F. Gell

Charles F. Gell, M.D., D.Sc. (Med.)
Scientific Director
NavSubMedRschLab

Reviewed and Approved by:

Joseph D Bloom.

Joseph D. Bloom, CDR, MC, USN
Director
NavSubMedRschLab

Approved and Released by:

J. E. Stark

J. E. Stark, CAPT, MC, USN
COMMANDING OFFICER
Naval Submarine Medical Center

SUMMARY PAGE

THE PROBLEM

To determine the influence of word predictability in sentence intelligibility tests which are used to evaluate speech discrimination ability in Navy and civilian personnel.

FINDINGS

It was concluded that word predictability is a factor which influences sentence intelligibility and that careful selection of key words on the basis of their predictability status could affect the overall intelligibility of sentences.

APPLICATION

The data may be incorporated into speech discrimination tests used to evaluate the hearing capabilities of Navy personnel. The results also apply to improvement of speech intelligibility among Navy divers during deep submergence operations and in other Navy environments where a degradation in speech intelligibility occurs.

ADMINISTRATIVE INFORMATION

This investigation was conducted as a part of Bureau of Medicine and Surgery Research Work Unit M4306.03-2020DAC5 — Evaluation of Underwater Communications Systems for Navy Divers and ONR Contract with the University of Connecticut N00014-67-A-0197-0001) and Mr. Duffy is a candidate for the Doctoral Degree and Dr. Giolas is Professor of Speech at the University of Connecticut. The present report is No. 6 on the BuMed Work Unit. It was approved for publication on 8 July 1971 and designated as Naval Submarine Medical Research Laboratory Report No. 672.

PUBLISHED BY THE NAVAL SUBMARINE MEDICAL CENTER

ABSTRACT

In line with the Naval Submarine Medical Research Laboratory's continuing effort to improve communication in the Navy, this study was instituted to investigate the relationship between word predictability and sentence intelligibility. This relationship was examined by comparing the accuracy of responses by listeners to several lists of sentences. Three methods of scoring used different groups of key words which had previously been judged to represent different degrees of predictability. It was hypothesized that the scores obtained would be a function of the predictability status of the key words used in scoring. Results indicated significant differences between the three scoring procedures for each sentence list under two filtering conditions, and these differences were in the hypothesized direction. The results suggest that the use of easy-to-predict words will increase sentence intelligibility while the use of difficult-to-predict words will depress intelligibility. It was concluded that word predictability is a factor influencing sentence intelligibility and that careful selection of key words, on the basis of predictability may be a way of controlling the intelligibility of sentences.

STUDIES IN NAVY COMMUNICATION: THE EFFECT OF WORD PREDICTABILITY ON SENTENCE INTELLIGIBILITY

INTRODUCTION

Monosyllabic word lists [1,2] have enjoyed widespread use in the assessment of speech discrimination ability, primarily due to the ease of including phonetic elements in a proportion comparable to their relative occurrence in normal conversational speech. Words most commonly used can be incorporated minimizing effects of vocabulary and listeners' intelligence. Furthermore, such tests are easy to administer and score. However, monosyllabic word lists do not adequately sample factors such as context, stress, accent, intonation voice quality and duration which normally provide cues to intelligibility in conversational speech. Some [2,3,4,5] have suggested that words embedded in sentences may be a more realistic measure of reception of conversational speech.

Although the use of sentences may overcome some of the disadvantages encountered with monosyllabic word lists, several characteristics of the average sentence should be investigated prior to recommending their general use. One such characteristic is word predictability, that is, the property of a sentence which permits prediction of a missing word(s) in that sentence. Because of varying contextual clues, the predictability status differs for words within a given sentence [6], as well as between sentences for the same word [7].

Since word predictability may play an important role in the intelligibility of connected speech, that role must logically be quantified before an effective test instrument can be constructed. This study fills that need.

PROCEDURE

A. Selection of Sentence Lists. Predictability values for several sentence lists were determined by Giolas, et al [6]. In that study, each list of sentences was recorded in its entirety and varying percentages of the total number of key words in each list were then eliminated by splicing out the randomly selected key words and replacing them with identical amounts of leader tape. Each group of subjects listened to the sentence lists under one of several word elimination conditions and were asked to write down what they felt were the missing words. Analysis of these results indicated that the C.I.D. Sentence Lists B and D [8] and the Revised C.I.D. Sentence List C [9] provided a wide range of predictability values for the key words within each list. Consequently, these lists were selected for use in the present study.

B. Preparation of Stimulus Material. Sections of the master tape of the prior study were used [6]. In order to increase error responses to avoid the "ceiling effect", we re-recorded the three sentence lists incorporating low-pass

filtering at 420 and 360 Hz using an Allison Model 2B filter (36 dB/oct). A pilot group of five listeners determined that these cut-off points yielded scores ranging between 30% and 70%. The VU meters of the playback and record units were matched during our re-recording, and a 1000-Hz calibration tone was inserted on each tape.

The practice sentences, five for each filtering, were inserted on the tape before the actual test sentences to minimize response errors on the initial test sentences due to the subjects' unfamiliarity with the novel listening task.

C. Subjects. Sixty Submarine School Candidates were used, two groups of 30 each. Each group was first given the taped, pure-tone audiometric screening test at 0.5 - 8 kHz. Those who failed (Hearing Level 25^{+} dB re ISO) two or more frequencies in either ear or at the same frequency in both ears were eliminated.

D. Presentation. The test tapes were played on an Ampex PR-10 tape recorder through an Altec 1569A amplifier to 49 matched TDH-39 earphones mounted in Otocups, in a room considered to be a good listening environment. Playback level was established by having two normal-hearing people listen to the tapes under test conditions and judge a comfortable level of loudness.

One group responded to monaural presentation of all three lists with the 420 Hz, the other, with the 360 Hz filter. Each group was given the following instructions:

"This is a study to see how well you can understand three groups of sentences which are distorted in a certain way. Each sentence will be preceded by its number. Then you will hear the sentence. It will be said only once, so listen carefully. Then I will stop the tape and you are to write down, word for word, the sentence you heard. If you are not sure, take a guess. Try to respond in complete sentences, and do your best not to leave any sentence blank. Write as neatly as possible and refrain from comparing your answers to those of others as this will affect test results. Before the test begins you will listen to five practice sentences which will give you an idea of the kind of distortion you will be listening to. Do not write these sentences down. After listening to them we will begin the test."

After listening to the five practice sentences, subjects were allowed to ask any questions. Experimental lists were then played in the order B, D, C, with a 2-3 min. rest period after each list.

E. Scoring. A subject's score with such lists is usually the number of all 50 key words in each list correctly identified. In this study, we also scored the number of correctly identified words from among the 20 most and 20 least predictable [6].

Homophonous words, as well as identifiable misspelled words, were accepted as correct. Contractions of words, as well as both words being spelled out, were also accepted as correct.

RESULTS AND DISCUSSION

All mean data are in Table 1. In order to analyze the relationship between scoring procedures, lists and filterings, a three-way analysis of variance was performed[10]. Since the

TABLE I

MEAN SCORES FOR EACH SCORING PROCEDURE FROM TWO GROUPS OF SUBJECTS LISTENING TO THREE SENTENCE LISTS UNDER TWO DIFFERENT FILTERING CONDITIONS (420 or 360Hz)

	C.I.D. List B			C.I.D. List D			Revised C.I.D. List C		
	FL*	EPL**	DPL***	FL*	EPL**	DPL***	FL*	EPL**	DPL***
420Hz low pass filtering	67.73%	75.83%	63.83%	67.66%	74%	59%	76.2%	84.5%	69.33%
360Hz low pass filtering	45.6%	48.5%	41.83%	34.46%	42.33%	25.33%	56.86%	65.66%	48.5%
combined means	56.66%	62.16%	52.83%	51.06%	58.16%	42.16%	66.53%	75.08%	58.91%

*FL = Scores based on 50 key words.
**EPL = Scores based on a list of the easiest to predict words.
***DPL = Scores based on a list of the most difficult to predict words.

scoring procedures used two different numbers of key words, all scores were first converted to percent correct. The results are in Table II.

The F obtained for the interaction of filtering, list, and scoring was not statistically significant. It was, therefore, appropriate to look at each factor independently, as well as its interaction with any one of the other two factors.

Bartlett's test for homogeneity of variance was performed and the non-significant results of this test offered no evidence that the variance across all conditions was not sufficiently homogeneous for further uncorrected analysis of results.

A. Differences Between Filter Conditions. As expected and as can be seen in Table II, a significant F was obtained

TABLE II

RESULTS OF THREE-WAY ANALYSIS OF VARIANCE OF THREE SCORING PROCEDURES USED TO EVALUATE RESPONSES FROM TWO GROUPS OF SUBJECTS TO C.I.D. SENTENCE LISTS B AND D AND REVISED C.I.D. LIST C UNDER TWO LOW PASS FILTERING CONDITIONS (420Hz AND 360Hz)

Source of Variation	Sum of Squares	df	Mean Square	F	p*
A (Filtering)	87401.7	1	87401.7	428	.01
B (Scoring Technique)	17224.7	2	8612.35	42.26	.01
C (Sentences)	24387.4	2	12193.7	59.84	.01
Interactions					
AB	25.2	2	12.6	.06	
AC	4084.8	2	2042.4	10.02	.01
BC	978.4	4	244.6	1.20	
ABC	317.6	4	79.4	.39	
within cell (experimental error)	106369.2	522	203.77		
Total	240789.0	539			

p* = point in the F distribution

for the two filtering conditions indicating that poorer scores were obtained with the more limiting filter condition.

B. Differences Between Sentence Lists. Significant F's (.01 level of confidence) were obtained for Lists B, D, and C (Table I), indicating that there truly are differences between these sentence lists. The significant interaction between filter settings and sentence lists (.01 level of confidence) further indicates that the two filterings had differential effects on the lists.

To probe the nature of the differences between the mean intelligibilities for each list, the Newman-Keuls method for computing critical differences was employed [10] (see Table III). The means collapsed over all scoring procedures for each list, between and within conditions, were significantly different at the .01 level of confidence, except for lists B and D in the 420 Hz low-pass condition; thus the responses to the different lists were not the same for the two filterings, and the question of list equivalency for Lists B and D

TABLE III

EVALUATION OF THE CRITICAL DIFFERENCES AS OUTLINED BY WINER (1962) OF THE DIFFERENCES BETWEEN SENTENCE LISTS BETWEEN AND WITHIN THE TWO FILTERING CONDITIONS USING THE COMBINED MEANS OF THE THREE SCORING PROCEDURES FOR COMPARISONS

		420Hz Low Pass			360Hz Low Pass		
		List B	List D	List C	List B	List D	List C
420Hz low pass	List B	____	2.25%	7.54%*	23.82%*	35.09%*	12.13%*
	List D	____	____	9.81%*	11.57%*	32.84%*	9.88%*
	List C	____	____	____	31.36%*	42.63%*	19.67%*
360Hz low pass	List B	____	____	____	____	11.27%*	11.69%*
	List D	____	____	____	____	____	22.96%*

* = significant at the .01 level of confidence

arises. Undoubtedly, the equivalency of both the original and the revised C.I.D. lists should be investigated.

C. Differences Between Scoring Procedures. Note in Table II the significant F's (.01 level of confidence) obtained for scoring procedures. The relationship between the mean scores for each scoring procedure (listed in Table I), under all test conditions, is illustrated in Figure 1. The results of a Neuman-Keuls test for critical differences indicate significant differences between scoring procedures for all lists and filterings and these differences are in the hypothesized direction. That is, the easy-to-predict words consistently yielded the highest scores, the difficult-to-predict, the lowest scores, and the full lists of 50 words yielded scores which consistently fell intermediate. The non-significant (see Table II) interactions of scoring/filtering and scoring/sentences further indicates that differences between scoring procedures were similar for all lists under filtering.

Although differences between the easy-to-predict and the difficult-to-predict scores were sometimes small, the smallest differences being 6.7% (List C, 360 Hz, low-pass filtering), and the differences were most often appreciable, (15% or greater in 4 of 6 instances, see Figure 1), the differences between the full list scoring procedure and the other two procedures was usually relatively small (differences ranged from 2.9% to 9.1%).

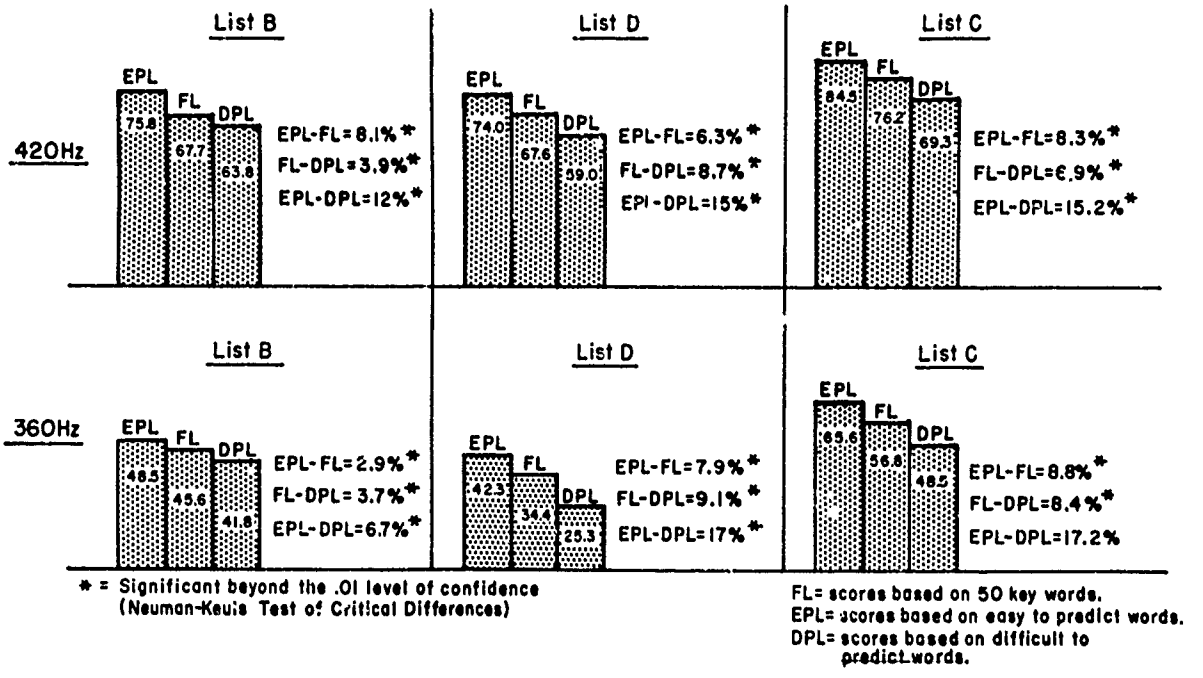

Fig. 1. Bar graphs illustrating the relationships between the means obtained for each scoring procedure within each sentence list over the two filtering conditions as well as the results of the critical differences evaluation for each of the scoring procedures within each sentence list and filtering condition.

It should be noted that when key words were eliminated in the study of Giolas, et al[6], they were completely eliminated, while in the present study all key words were present in the message. The influence of additional acoustic cues in the present investigation is unknown, but it is certainly possible that the acoustic cues have altered the predictability status of the key words, in the sense that some key words were acoustically more intelligible than others. However, any systematic bias toward one particular scoring procedure seems quite unlikely.

When scores for these lists are compared to isolated word intelligibility [11,12], it is quite apparent that under several low-pass filterings, sentences are considerably easier to understand than words. Of course context is well known to influence speech intelligibility.

CONCLUSIONS

This study demonstrates a close relationship between the predictability of words and the intelligibility of sentences incorporating those words. The data show intelligibility of sentences can be partially controlled by selecting key words for predictability. The results suggest that if isolation of parameters affecting intelligibility is desirable, any further development of sentence lists for use as tests of speech intelligibility carefully consider the relative predictability of the key words.

These data reveal that predictability is an important factor influencing sentence intelligibility tests such as used currently and probably in the future to evaluate speech reception. These results can be incorporated into further refinements of sentence tests designed to evaluate speech reception in Navy personnel. The results can be applied to improving speech communications among Navy personnel working in environments where degredation in speech intelligibility exists.

REFERENCES

1. Egan, J. P., Articulation Testing Methods, Laryng. 1948, 58, 955-991.

2. Hirsh, I. J., Davis, H., Silverman, S. R., Eldert, E. G. and Benson, R. W., Development of Materials for Speech Audiometry, J. Speech Hear. Disord. 1952, 17, 321-337.

3. Lehiste, I. and Peterson, G. E., Linguistic Considerations in the Study of Speech Intelligibility. J. Acoust. Soc. Am. 1959, 31, 280-286.

4. Speaks, C. and Jerger, J., Method for Measurement of Speech Identification, J. Speech Hear. Res. 1965, 8, 184-194.

5. Giolas, T. G., Comparative Intelligibility Scores of Sentence Lists and Continuous Discourse, J. Aud. Res. 1966, 6, 31-38.

6. Giolas, T. G., Cooker, H. and Duffy, J., Word Predictability Employing a Modified Cloze Procedure, Unpublished manuscript, 1969.

7. Howes, D., The Intelligibility of Spoken Messages, Am. J. Psychol. 1952, 65, 460-465.

8. Davis, H. and Silverman, S. R. (Ed), Hearing and Deafness, Revised New York: Holt, Rinehart and Winston, 1960.

9. Harris, J. D. et al, The Relation Between Speech Intelligibility and the Electroacoustic Characteristics of Low Fidelity Circuitry, J. Aud. Res. 1961, 1, 357-381.

10. Winer, B. J. Statistical Principles in Experimental Design, New York: McGraw-Hill, 1962.

11. Hirsh, I. J., Reynolds, E. and Joseph, M., Intelligibility of Different Speech Materials, J. Acoust. Soc. Am. 1954, 26, 4.

12. Epstein, R., Giolas, T. G. and Owens, E., Familiarity and Intelligibility of Monosyllabic Word Lists, J. Speech Hear. Res. 1968, 11, 435-438.